

메타데이터언어 및 저장체계의 개발 방법에 관한 연구

A Study on the Development Methods of Metadata Language
and Repository Architecture

박동진*

목 차

- | | |
|-------------------|--------------------------|
| I. 서론 | III. 메타데이터 기반 시스템 개발 방법론 |
| II. 연구의 배경 | 1. 메타데이터언어 개발 방법 |
| 1. 메타데이터 언어 | 2. 저장체계의 개발 방법 |
| 2. 메타데이터 기반 정보시스템 | IV. 결론 |

Key Words : 메타데이터, 메타데이터언어, XML 스키마, 개발방법론

Abstract

아직 국내에서는 표준 메타데이터를 기반으로 한 정보시스템 개발의 경험이 없으며 최신의 메타데이터 개발도구, 절차, 핵심기술 및 통합방법에 관한 가이드라인이 없다. 그러나 향후 3-4년 이후에 본격적인 메타데이터 시스템의 개발이 시작될 예정인데 각 기관에서는 미래의 시스템에 대하여 어떻게 준비를 하고 진행을 해야 할지를 모르고 있는 실정이다. 따라서 현재 외국에서 성공적으로 개발되어 운영되고 있는 메타데이터 기반 시스템을 벤치마킹하여 시스템 개발프로세스를 표준화하고, 이를 기반으로 한 시스템의 개발 방법론에 대한 연구가 시급한 상황이다. 본 연구는 이를 위한 사전 연구에 해당하는 것으로 구체적인 연구의 목표는 메타데이터 언어와 저장체계 개발과 관련된 프레임워크를 제시하는 것이다.

I. 서론

최근 각 분야에서 메타데이터를 중심으로 한 정보통합 관리기술의 채택이 고려되고 있으며, 국가적으로 그리고 각 기관에서는 차세대 정보관리시스템에 메타데이터 기술들을 적용할 예정이다. 이러한 차원에서 많

은 기관에서는 메타데이터 표준을 제정하고 있다. 그러나 현재의 표준은 문서상의 표준일 뿐 각 정보관리시스템에 적용할 수 있는 시스템으로의 표준은 아니다. 따라서 현재의 표준을 활용하여 정보시스템에서 표준 메타데이터를 생성하고 이를 통합하는 기술을 개발하고 이를 검증하는 것은 현 시점에

* 공주대학교 산업시스템공학과

서는 매우 중요하다.

그리고 아직 국내에서는 표준 메타데이터를 기반으로 한 정보시스템 개발의 경험이 없으며 최신의 메타데이터 개발도구, 절차, 핵심기술의 파악 및 통합방법에 관한 가이드라인이 없기 때문에, 향후 3-4년 이후에 본격적인 메타데이터 시스템의 개발이 시작될 예정이나 각 기관에서는 미래의 시스템에 대하여 어떻게 준비를 하고 진행을 해야 할지를 모르고 있는 실정이다.

본 연구는 현재 외국에서 성공적으로 개발되어 운영되고 있는 메타데이터 기반 시스템을 벤치마킹하여 시스템 개발프로세스를 표준화하고, 이를 기반으로 정보시스템의 개발 방법론을 구체적으로 제시하기 위한 사전연구에 해당된다. 구체적인 연구의 목표는 크게 메타데이터 언어와 저장체계 개발과 관련된 프레임워크를 제시하는 것이다. 연구의 전체적인 방향은 다음과 같다. 첫째, 각 기관에서 적용가능한 본원적 모델(generic model)을 제시한다. 둘째, 상호운용성 확보를 위하여 국제/국내의 표준스키마를 활용한다. 셋째, 연구성과정보를 코드화하는 포맷으로 XML을 선택한다. 제2장에서는 연구의 배경을, 제3장에서 메타데이터 시스템 개발방법을 제시하고 제4장에서 결론을 맺는다.

II. 연구의 배경

1. 메타데이터 언어

현재까지 메타데이터언어에 대한 통일된 정의는 없다. 일부 논문에서 논의 되는 것처럼(Ramachandran et al., 2001; McCartney and Jones, 2002), 본 연구에서도 메타데이터언어란 데이터의 상호운용성 문제를 해결(solution)하는 통합정보기술로 본다. 그리고 메타데이터 언어의 개발에는

다음과 같은 원칙이 있다. 첫째, 메타데이터언어는 업계에서 범용으로 사용되는 기계가 파싱가능한 포맷이 되어야 하며 특정 플랫폼 및 소프트웨어와 독립적이어야 한다. 둘째, 호환성을 향상시키고 중복성을 제거하기 위해서는 기존의 가용한 메타데이터 표준을 활용한다. 셋째, 표준은 개별사이트에서 메타데이터를 만들고, 저장하고, 관리하는데 사용되기보다는 서로를 통합하는데 그 주안점을 두어야 한다(McCartney and Jones, 2002). 메타데이터언어의 구성요소로는 메타데이터 문서(file)와 이 문서를 정의한 XML 스키마, 그리고 각종 툴들을 포함하는 소프트웨어 라이브러리 등이 포함된다.

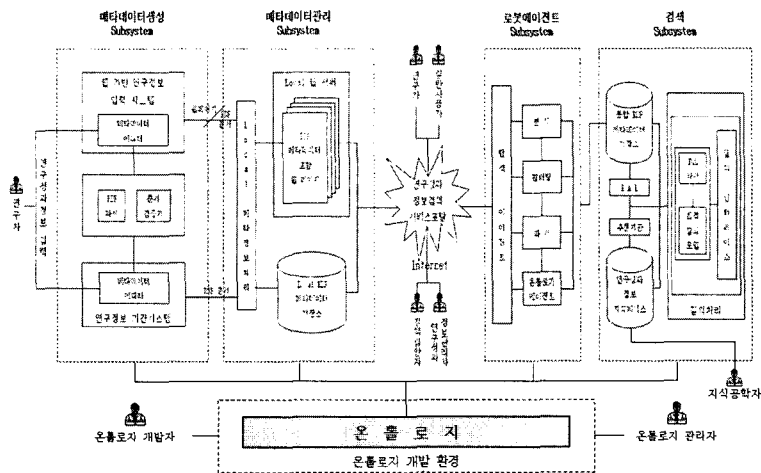
예를 들면 (미)알라바마 대학교에서 지구과학분야의 메타데이터 관리를 위한 메타데이터언어인 ESML(Earth Science Markup Language)을 개발하였다(Ramachandran et al., 2001). 사용자는 ESML 에디터를 이용하여 ESML 스키마 규칙 안에서 ESML 메타데이터 파일을 생성한다. ESML 라이브러리에는 다양한 기능을 하는 응용프로그램들이 포함된다. 라이브러리의 핵심기능으로는 ESML 파일을 읽고, 파일로부터 데이터들을 발췌하는 것이다. 이때 파일을 DOM(Document Object Model)으로 변화시키는 파싱 프로그램이 라이브러리에 포함된다. 그리고 메타데이터의 저장 및 부라우징을 위한 모듈들도 포함된다. 각 로컬시스템의 개발자는 ESML 라이브러리의 API를 이용하여 시스템을 개발한다. 또 다른 메타데이터언어의 예로 미국 NCEAS와 LTER에서 개발이 주도된 EML(Ecological Metadata Language)이 있다(Higgins et al., 2002; Fegraus et al., 2005). EML은 생태자료의 메타데이터를 기술하는 정규화되고 표준화된 방법을 제시한다.

2. 메타데이터 기반 정보시스템

메타데이터 개발단계는 개발의 목적, 범위 등에 따라서 많이 다르므로 일반적으로 통용되는 절차 및 방법은 없다. 대만에서 디지털 도서관 및 박물관을 위한 메타데이터 개발에서 채택한 절차와 내용을 제시하였다(Chen and Chen, 2001). 또한 Xie와 Shibasaki(2005)에 의해서 지구상의 Water Cycle과 관련된 각종 자료 관리 및 서비스를 위한 메타데이터 시스템이 개발되었다. 이 시스템은 크게 3개의 과정을 거쳐서 개발되었다. 첫째는, 메타데이터개념 모델의 구축이다. UML으로 기술한 이 모델은 메타데이터의 요소를 핵심(공통)요소, 분야별 특성요소, 그리고 조직관련 요소 등으로 구성하였다. 둘째는, 메타데이터 구현 모델로서 메타데이터개념 모델을 XML Schema, XSL

로 구현을 하였다. 마지막으로 웹서비스 응용시스템을 개발하였다. 이는 메타데이터의 저장 및 관리와 웹을 통하여 이용자에게 각종서비스를 제공할 수 있는 기능을 포함한다.

박동진(2005)은 이상적인 메타데이터 기반 연구성과정보시스템의 개념적 구조를 아래 <그림 1>과 같이 제시하였다. <그림 1>에서처럼 온톨로지를 기반인프라로 하며 기능적으로 크게 4개의 서브시스템으로 구성된다. 즉 메타데이터생성 서브시스템과 메타데이터관리 서브시스템은 각 연구기관에서 관리되며, 로봇에이전트 서브시스템과 검색 서브시스템은 국가수준에서 연구성과정보를 서비스하는 기관에서 관리한다. 현재 개발되어 사용 중인 국외의 메타데이터 정보시스템을 정리하면 아래와 같다.



<그림 1> 메타데이터 기반 연구성과물 관리시스템 개념적 구조

CERIF-2000을 기반으로 Lopatenko와 동료들에 의해서 유럽의 연구기관간의 연구정보 교환을 위해 연구성과정보시스템 프로토타입인 AURIS-MM이 개발되었다(Lopatenko, 2001). 이것은 기존의 오스트리아연구정보시스템인 AURIS에 시맨틱 웹기술을 채택하여 기능적으로 확장한 시스템이다. 이 시스템은 CERIF-2000을 기반으로 온톨로지를 개

발하고 이를 RDF/RDFS 로 표현하였다. 또한 IST World Web Portal은 유럽 내의 국가단위의 연구센터(RTD: Research and Technology Development)의 정보를 통합한 서비스 포털이다(Jorg, 2006). 이 시스템의 메타데이터 저장체계는 CERIF-2005 Full Data Model을 기반으로 설계하였다.

미국이 중심이 된 KNB(Knowledge Network

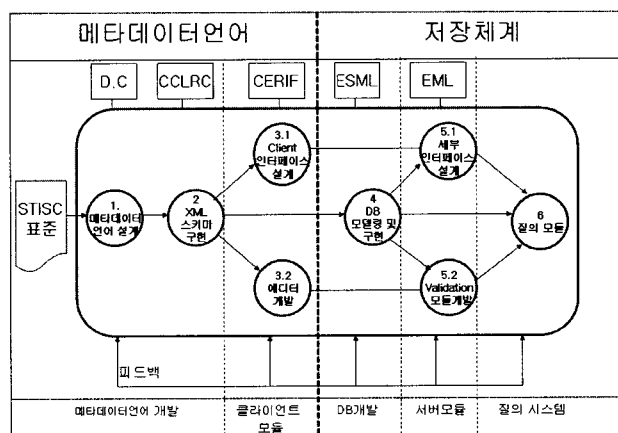
for Biocomplexity)는 생태학 분야의 국제 연구 네트워크이다. 여기에서는 생태 및 환경정보를 공유하기 위하여 EML을 표준스키마를 이용한 시스템을 개발하여 활용하였다 (Lin, 2006). Morpho는 생태학자 스스로 메타데이터를 생성, 질의, 편집이 가능하도록 하는 사용자를 위한 통합 툴이다. Metacat은 XML 메타데이터 및 데이터의 저장체계 (repository)로서 EML에 최적화 되어있다 (Jones et al., 2001). 그리고 관계형 DB에 저장된 메타데이터를 관리하는 Xanthoria 등이 있다. 이것은 데이터베이스 연결을 가능하게 하는 응용프로그램으로서 데이터를 발췌하고 포맷화하는데 스타일 시트를 이용한다. 특히 Xanthoria는 EML 2.0을 위한 솔루션이다.

e-사이언스의 주도 그룹인 CCLRC에서는 다양한 메타데이터 시스템을 개발하여 운영하고 있다. ICAT(ISIS Metadata Catalogue)는 CCLRC 메타모델을 기반으로 한 메타데이터 카탈로그 시스템으로 연구프로젝트 정보의 저장시스템과 같은 역할을 한다. 즉 연구프로포잘, 보고서, 원시데이터, 분석데이터, 각종 연구성과물의 메타데이터를 저장, 검색 및 관리할 수 있는 시스템이다 (Flannery, 2004). CCLRC Data Portal은 CCLRC에서 운영하고 있다(Drinkwater, 2004). Data Portal이란 다양한 과학데이

터를 하나의 인터페이스를 통하여 접근할 수 있는 게이트웨이의 역할을 한다. 또한 CCLTC e-Science 센터에서 서비스하는 각종 GRID 기반의 서비스에 접근할 수 있는 통로의 역할도 한다. Data Portal은 분야별 ICAT의 메타데이터 카탈로그 시스템과 표준 웹 데이터페이스로 통합된다.

III. 메타데이터 기반 시스템 개발 방법론

메타데이터 기반 시스템의 개발절차를 보면 먼저 기존의 메타데이터 표준을 바탕으로 메타데이터언어를 설계한다. 설계된 언어를 XML 스키마로 변환시킨 다음에, 스키마의 룰에 적합한 문서를 산출하는 메타데이터언어 에디터를 개발한다. 다음으로 메타데이터언어에 적합한 스키마 기반의 데이터베이스를 개발한다. 그리고 서버 환경에서 데이터베이스를 활용하는 기능들을 S/W화 하고, 마지막으로 메타데이터 질의시스템을 개발한다. 각 시스템 모듈들은 컴포넌트로 설계하여 구현할 수 있다. <그림 2>는 본 연구의 상세프로세스를 보여준다. 그리고 각 단계의 개발내용을 구체적으로 설명하면 아래와 같다.



<그림 2> 메타데이터 시스템 개발 프로세스

1. 메타데이터언어 개발 방법

메타데이터언어는 대상이 되는 정보를 기술하는 핵심 개념들을 정규화(formalizing)하고 표준화(standardizing)한 것이다. 메타데이터언어는 다음과 같은 원칙 하에서 개발한다. 첫째, 기계가 구조를 인지할 수 있는 형태(machine parseable format)로 코드화 하며 특정 플랫폼 및 소프트웨어와는 독립적이어야 한다. 둘째, 호환성(compatibility)을 확보하고 중복성(redundancy)을 최소화하기 위하여 기존의 국내외 메타데이터 표준들을 충분히 고려하여야 한다. 셋째, 하나의 사이트에서 메타데이터를 만들고 저장하고 관리하는 도구뿐 아니라 분산되고 이질적인 데이터를 통합하는데도 초점이 맞추어져야 한다.

메타데이터문서를 코드화 하는 포맷으로 XML을 선택한다. XML은 문서의 내용을 태깅하고 규격화된 스키마로 문서를 검증할(validating) 수 있다는 장점이 있으며, XSL을 이용하여 다른 포맷으로 문서를 변형시킬 수도 있다. 또한 분산/이질적인 개별 사이트 간에 상호운용성을 확보하는 기반 기술로서 XML의 채택이 바람직하다. 세부 사항을 보면 다음과 같다.

1) 메타데이터언어 설계 및 명세

메타데이터언어는 기본적으로 Dublin Core 의 “자원(resource)” 개념과 ISO로부터 식별데이터요소를 기반으로 설계한다. 특정 자원에 대한 정보는 일련의 데이터요소의 집합으로 규정한다. 메타데이터언어는 하나의 큰 시스템 표준이 아니라 여러 모듈들의 집합으로 설계하는데, 각 모듈은 물리적 대상 정보를 논리적인 내용으로 구분하는 것이다. 모듈은 필요에 따라서 넓이 그리고 깊이(breadth and depth) 측면에서의 확장이 가능해야 하며, 각 기관에서는 그들이 데이터를 기술하고자 대상에 따라 적절

한 모듈을 선택하고 모듈을 수정 및 보완할 수 있어야 한다.

2) XML 스키마 및 파서

전 단계에서 개발된 메타데이터언어 명세서는 텍스트로 작성된 문서이다. XML 스키마는 앞에서 설계된 모듈의 계층적 구조로 트리와 하위트리로 구성된다. 먼저 스키마의 구조는 XML Spy 최신버전을 사용하여 스키마 다이어그램을 작성한다. 계층화된 XML 스키마는 메타데이터 문서를 작성하고 유효성을 검증하는데 활용된다. 다음으로 메타데이터 문서가 XML 스키마에 적합하게 작성되었는지에 대하여 검증(validation)하기 위하여 온라인 파서를 개발한다. Java 기반에서 DOM 혹은 SAX를 활용하여 파서를 개발할 수 있다.

3) 클라이언트 모듈

연구자가 XML 스키마를 참조하여 직접 메타데이터를 생산하는 것은 불가능하다. 즉 이를 위해서는 사용자 입장에서 편리하게 사용할 수 있는 응용프로그램이 필수적이다. 이 응용프로그램은 메타데이터 저장체계와 인터페이스하면서 다음과 같은 역할을 해야 한다. 첫째, XML 문법을 적용하여 메타데이터를 생성하고 편집하는 에디터 기능이 필요하다. 둘째, 메타데이터 입력품의 생성단계를 도와주는 지능적 인터페이스 역할을 해야 한다. 셋째, 데이터베이스로부터 검색된 결과를 다양한 포맷(HTML 및 XML)으로 디스플레이하는 기능이 필요하다.

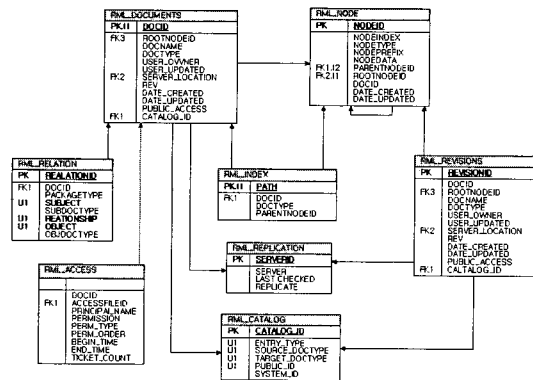
2. 저장체계의 개발 방법

표준 메타데이터 언어를 이용하여 상호운용성을 확보하기 위해서는 분산되어 있는 XML 데이터를 저장하기 위한 표준 저장체계가 개발되어야 한다. 생성된 XML 데이터의

저장은 XML 전용 데이터베이스와 관계형 데이터베이스에 저장될 수 있으나, 현재 기술 수준으로는 XML 문서를 관계형데이터베이스에 매핑하여 저장하는 것이 현실적이다. 또한 서버수준에서 데이터베이스를 활용하는 기능들이 필요하며 마지막으로 질의시스템이 개발되어야 한다.

1) DB 설계 및 구현

XML 데이터의 구조를 모델링 하는 효과적인



<그림 3> 데이터 모델

2) 서버 모듈

저장시스템 서버의 프로토타입 구현은 크게 클라이언트 API, 자바 서블릿, 서버의 주요기능(메타데이터저장, 수정, validation, transformation)을 포함한다. 서버에는 클라이언트 시스템과의 인터페이스를 위한 클라이언트 API가 있어야 한다. 서버의 주요기능 중 유효성검사 기능은 XML 스키마의 검증 룰에 따라서 각종 메타데이터 문서를 검증하는 기능이다. 사용자가 XML 문서를 만들어서 저장할 때, 그리고 데이터베이스로부터 XML 문서를 발췌하였을 때 이 문서가 유효한 것인가를 확인해 주는 것이다. 특히 각 문서의 Public Identifier를 검사하고 부여하는 것도 중요한 역할이다. 레코드 형태로 데이터베이스에 저장된

인 방법이 DOM(Document Object Model)이다. 본 연구에서는 DOM을 이용하여 XML 데이터를 관계형데이터베이스에 저장하는 방법을 제시한다. DOM은 XML 문서를 계층적 논리적 구조로 정의하고 이를 일반 프로그래밍언어로 인터페이스가 가능하다. 따라서 프로그램으로 문서구조를 내비게이션하고, 엘레먼트나 내용을 추가, 수정, 그리고 삭제하는 것이 가능하다. DB 설계를 위한 논리적인 모델은 아래와 <그림 3>과 같다.

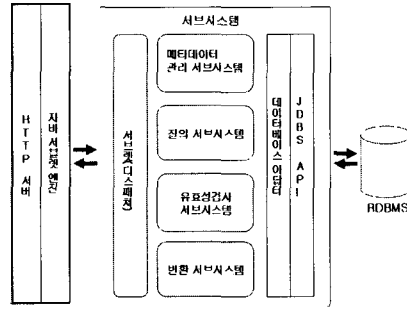
XML 문서가 발췌되고 난 다음에는 자동적으로 XML 문서 혹은 최종 사용자를 위하여 HTML 형태의 문서로 자동으로 변환되어야 한다. 이때 XSLT가 메타데이터를 효율적으로 표현 가능하게 한다.

3) 질의 시스템

서버시스템은 데이터베이스에 저장된 메타데이터를 검색할 수 있는 인터페이스를 제공하여야 한다. 검색은 다음과 같은 절차를 거친다. HTML으로 작성된 사용자의 검색 품으로부터 검색 요청된 내용이 서블릿 파라메타로 전달되면서 검색을 위한 pathquery 문서가 작성된다. 이때 pathquery를 자동으로 작성하는 해당 pathquery를 SQL 명령문으로 전환해주는 컴

포넌트의 개발이 필요하다. 그리고 데이터베이스에 적용된 SQL 명령문의 결과물인 레코드를 XML 문서로 자동으로 변환해 주어야 한다. 또한 최종사용자가 XML 문서를 직접 읽는 것은 의미가 없기 때문에 이를 특정 품으로 자동 변환해 주어야 한다. 즉 XSLT

를 적용시켜 HTML 형태로 사용자에게 전달할 필요가 있다. 본 질의 시스템 구현은 Xpath 와 XQuery 등의 기술상의 진보 및 표준화와 툴의 발전으로 인하여 다른 방법으로 구현할 가능성도 있다. 아래 <그림 4>는 2년차에 개발될 시스템의 개념적 구조이다.



<그림 4> 서버시스템 구성

IV. 결론

현재 국내에서 많은 단체들이 메타데이터 표준을 제정하였지만, 이것을 정보시스템 차원에서 적용하기 위한 시도는 거의 없다. 이는 아직 해당기술들이 안정적이지 못하고 계속적으로 진화되어 가고 있는 상태이며 시스템개발을 위한 검증된 방법론이 없기 때문이다. 본 연구는 표준의 제정에서 시스템 개발에 이르기까지 체계적이고 일관성이 있는 방법론을 개발하기 위한 사전연구이다. 본 연구는 아직 시작 단계이나 계속적인 연구에 의한 결과물은 메타데이터 정보

시스템의 개발 계획에 있어서, 그리고 개발 실무자들에게 현실적인 해결방안 및 지침으로서 역할을 할 것이다. 본 연구는 문제해결관점에서 다양한 학문분야의 최신기술, 방법론, 그리고 툴들을 적절히 통합하는 것이다. 즉 표준메타데이터를 활용한 데이터의 공유 및 서비스문제 해결이라는 초점에서 학문들 간의 통합과 연구영역을 구체적으로 파악하고 문제해결방안을 모색한다. 따라서 본 연구는 새로운 융합학문분야를 파악하고 지속적인 후속 연구를 가능하게 할 것이다. 따라서 현재 타 선진국가에 비해서 상대적으로 낙후된 이 분야의 연구를 촉진시킬 가능성이 있다.

참 고 문 헌

1. 박동진, 이상태, 최기석, “메타데이터 기반의 검색시스템의 개념적 설계,” *정보관리연구*, Vol. 37, No. 2, 2006
2. Chen, H and C. Chen, Metadata Development for digital Libraries and Museums -Taiwan’s Experience, DC-2001, NII, Tokyo, 2001.
3. Drinkwater G. and S. Sufi, The CCLRC Data Portal, Proc. UK e-Science Programme All Hands Meeting (AHM2004), Nottingham, UK, 2004.
4. Fegraus E, S. Andelman, M. Jones and M. Schildhauer, Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. *Bulletin of the Ecological Society of America*. Vol. 86, No. 3, 2005.
5. Flannery D., ISIS Data, Metadata, Proposals and the CCLRC Data Portal <<http://ins00.psi.ch/nobugs2004/papers/paper00147.pdf>>, 2004.
6. Higgins D., C. Berkley and M. Jones, Managing heterogeneous ecological data using Morpho, *Proceedings of 14th International Conference on Scientific and Statistical Database Management*. Edinburgh, UK. July 2002.
7. Jorg B., J Ferlez and E. Grabczewski, IST World: European RTD Information and Service Portal, *8th International Conference on Current Research Information Systems: Enabling Interaction and Quality: Beyond the Hanseatic League (CRIS 2006)*, Bergen, Norway, 11-13 May 2006
8. Lin C., H. John and P. Lu, A Metadata-based Framework for Multilingual Ecological Information Management, *Taiwan Journal for Science*, Vol. 21, No 3, 2006.
9. Lopatenko, A., A. Asserson, G. Keith, and K. Jeffery, CERIF - Information Retrieval of Research Information in a Distributed Heterogeneous Environment, *Proceeding on the 6th International Conference on Current Research Information System*, Aug. 2002.
10. McCartney, P. and M. B. Jones, Using XML-encoded Metadata as a Basis for Advanced Information Systems for Ecological Research. *Proceedings of the 6th World*, 2002.
11. Ramachandran R., M. Alshayeb, B. Beaumont, H. Conover, S. Graves, X. Li, S. Movva, A. McDowell and M. Smith, Earth Science Markup Language: A Solution for Generic Access to Heterogeneous Data Sets, *Earth Science Technology conference*, Maryland, August 28-30, 2001.
12. Xie, R. and Shibaski R, Standardization Framework for CEOP Metadata Development and Application, *CEOP/IGWCO Joint Meeting*, University of Tokyo, 2005.